Large-scale significance testing of the full Moon effect on deliveries

David Causeur*and Arthur Charpentier†

March 12, 2009

Abstract

As mentioned by many authors, the belief that the number of women going into labor and giving birth is higher during the full Moon is widespread, even among the medical staff. However, various statistical studies of the daily number of births along the Moon cycle, mostly on rather short periods (from 40 to 60 lunar cycles, i.e. less than 5 years), conclude to contradictory results, which strengthens the need for a powerful analysis on a large amount of data. We propose a large-scale significance testing analysis of the full Moon effect in each lunar cycles from 1968 until 2005 based on the daily numbers of births in France. A multiple testing methodology (see [13] and [7]) which accounts for dependence among lunar cycles is used to guarantee both a high overall power and a control of the False Discovery Rate at a low level. Results confirm the existence of a small yet marked full Moon effect: on average, one cycle per year shows a significantly larger birth rate during a 6-days period around the Full moon day than the other days of the cycle, which is four times more than for a comparison between any other 6-days period and the rest of the cycle.

Keywords: Full Moon effect; High-dimensional data; Multiple testing; Number of births.

1 Introduction

There is a persistent myth in several societies about the influence of lunar cycles on deliveries. Even nowadays, many delivery nurses attest that the number of women going into labor and giving birth is higher during the full Moon. For example, [5] mention that in 1987, 80% of nurses and 64% of

*Applied Mathematics Department, Agrocampus, CS 84215, 65 rue de Saint-Brieuc, F-35042 Rennes Cedex, `david.causeur@agrocampus-ouest.fr`.

†Université Rennes 1-CREM & CREST, Place Hoche, F-35000 Rennes, `arthur.charpentier@univ-rennes1.fr`.

doctors believed that there was an impact of the Moon Fifty years ago, this phenomenon has been termed "*the full Moon effect*", from [26].

Numerous studies tried to highlight this full Moon effect. This literature introduced early the idea of defining properly a window around the full Moon, i.e. a period of days before, during and after the full Moon over which a supposedly larger delivery rate may be expected. Some studies, e.g. [18], exhibited that, on average, highest birth rates occur around days with a full Moon. Most notably, other studies show that there were peaks during full Moon, and drops before and after, which strengthens the feeling of a full Moon effect on births (see e.g. [19] or [16]). Later, the idea of studying p-values of statistical tests has been mentioned as in [21], but here, no significant difference among birth rates has been observed. Several other ideas have been introduced to study the impact of the Moon cycle on birth rates, e.g. cross-correlation (see [4]), time series models (see [27]), or spectral analysis (see [10]).

Recently, studies in various countries have been performed, most of them on a rather short period (from 40 to 60 lunar cycles, i.e. less than 5 years). Results were contradictory, since either no correlation were found between lunar cycles and birth rates ([2], [22], [8], [11], [15], [20], [31]), or a significant one has been measured ([30], [1], [9]). A recent survey on five countries, in [23] mentions that there should be no evidence of lunar impact on deliveries. These different conclusions suggest that the demonstration of an eventual full Moon effect should necessitate both a powerful statistical methodology and a large amount of data.

In the present paper, a statistical study of the daily numbers of births in France from 1968 until 2005 is proposed, focusing on multiple comparisons of the mean numbers of deliveries during and after the full Moon period in each lunar cycle. Most multiple testing procedures are based on the following principle: a null hypothesis is rejected if the corresponding p-value does not exceed a threshold τ, which value is chosen so that a pre-chosen type-I error rate is controlled at a low level. Multiple testing issues has long been considered almost exclusively in Analysis of Variance settings where adjustment for the multiplicity of tests is required to control the probability of at least one false rejection, also called Family-Wise Error Rate (FWER), when testing linear contrasts simultaneously. Generally in this context, the limited number of contrasts and the duality between multiple testing and simultaneous confidence interval has almost exclusively motivated the development of FWER-controlling multiple testing procedures. However, these methods can result in too conservative decision rules when applied to high-dimensional data, such as in the present situation, where the large number of lunar cycles regarding the number of observations per cycle and the potentially high amount of dependence due to the temporality of the data also discourages from using FWER-controlling procedures.

The seminal paper by [3], introducing the False Discovery Rate (FDR) as an overall type-I error rate in multiple testing for high-dimensional data, has substantially renewed the methodology. This type-I error rate is defined as the expected proportion of erroneous rejections among the rejected hypotheses. FDR-controlling procedures have indeed shown desirable properties when applied to high-dimensional data, often resulting in less conservative decision rules. In the last two decades, many improvements of the initial Benjamini-Hochberg (BH) procedure has been proposed, mainly for a more strict control of the FDR under various dependence assumptions. Many authors have also pointed out the negative impact of a large amount of dependence both on the stability of the error rates and on the power of the multiple testing procedure (see [6], [13]). Among the most recent ideas to reduce this negative impact, [13] and [7] have both suggested to account for dependence between the test statistics by means of a factor modelling of the intra-group variance. This multiple testing methodology is used to find out lunar cycles with a significant full Moon effect, taking advantage of a factor structure for the intra-period variance of the number of births. Section 2 is dedicated to a large-scale exploratory analysis of the daily numbers of births and a pre-processing of the data to correct for obvious trends and patterns and to remove outliers. In section 3, the factor-analytic procedure mentioned above for multiple testing is used to point out moon cycles with significant full Moon effect. Finally, section 4 is dedicated to a discussion of the results.

2 Removing trends and patterns in the raw series

The data are obtained from the INSEE (National Institute for Statistics and Economics, in France) and contain daily number of births in France from January 1st, 1968, until December 31st, 2005, i.e. 29,385,552 births (Figure 1), denoted X_t. For convenience, we assume that deliveries on a given day occur at noon. The Moon cycle has a period of 29 days, 12 hours and 44 minutes (a fixed arithmetic lunar calendar was considered, rather closed to the Orthodox Easter computus or the Hebrew calendar molad, at least on a 40 year basis). The starting point of the cycle was the first full moon of the data set, observed[1] at 16:11 UTC/GMT on January 15th, 1968 (see [17]).

Correction for the week-end effect

Figure 1 shows a week-end effect consisting in a growing difference along time between the mean number of births during week days and week ends. Therefore, the first step was to introduce a correction for this effect. A smooth pattern can be obtained by considering the ratio week-days/week-end (Figure 2).

[1]http://www.timeanddate.com/calendar/Moonphases.html?year=1968

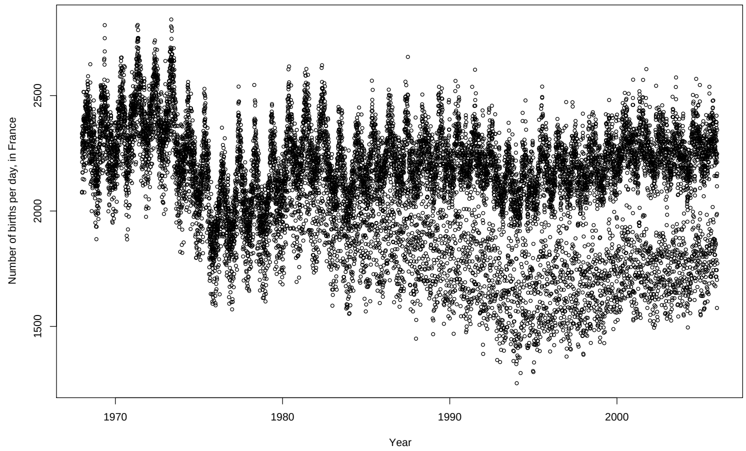

Figure 1: Number of births per day, (X_t), in France, from January 1st, 1968, until December 31st, 2005, i.e. 29,385,552 births.

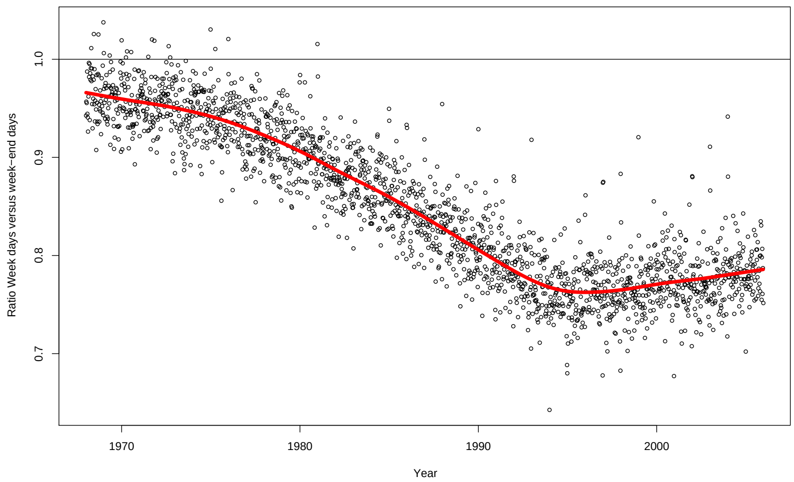

Figure 2: Weekly ratio of the daily number of births during the week-end and during the week before, from January 1st, 1968, until December 31st, 2005.

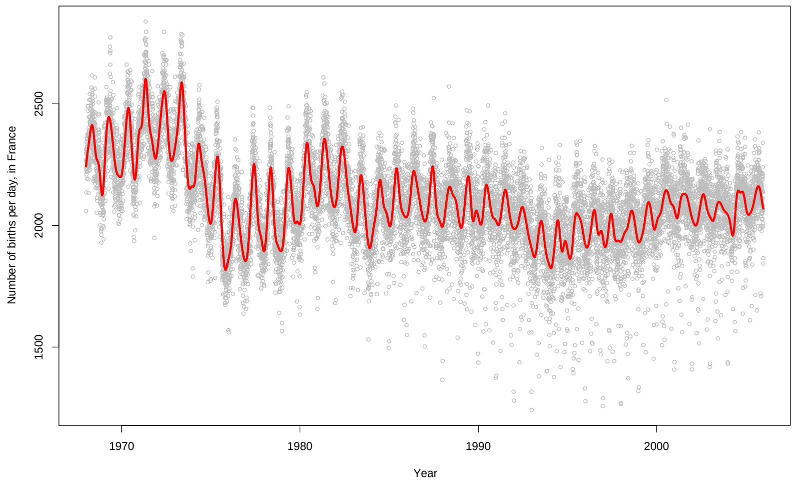

Figure 3: General (yearly) pattern in the series corrected from the weekly effect, Y_t.

Let a_t denotes this (smoothed) weekly ratio, we propose to define the weights for week days and week-end days by respectively $7a_t/(5a_t+2)$ and $7/(5+2a_t)$, i.e.

$$Y_t = \frac{7a_t}{5a_t+2}X_t \ \mathbf{1}(t \text{ is a week day}) + \frac{7}{5+2a_t}X_t \ \mathbf{1}(t \text{ is a week-end day}).$$

After this weekly pattern correction, a more general pattern has been considered on Y_t (see Figure 3), taking into account a general long term trend, and an annual cycle (since there are more births in May and less in September and October, for instance). This general pattern has been estimated using a kernel regression method (see e.g. [28]).

The focus of our study is the remaining noise of the initial series, removing the general pattern (see Figure 4).

Outliers' detection

Figure 4 shows a small number of data points with abnormally low values, especially at the end of the series, which suggests an undesirable effect of the above adjustment from general patterns. A robust regression method is now used to identify these outliers. A linear trend is locally fitted, using least-median of squares regression (see [24]), and points are considered as outliers if the corresponding absolute standardized residuals with respect to this robust fit exceeds 3.5. Note that neighborhoods over which the local fits are derived are large enough (made of 700 points) to ensure a smooth estimated regression function. Due to the rather conservative choice of a high critical value for the standardized residuals, a restricted number of

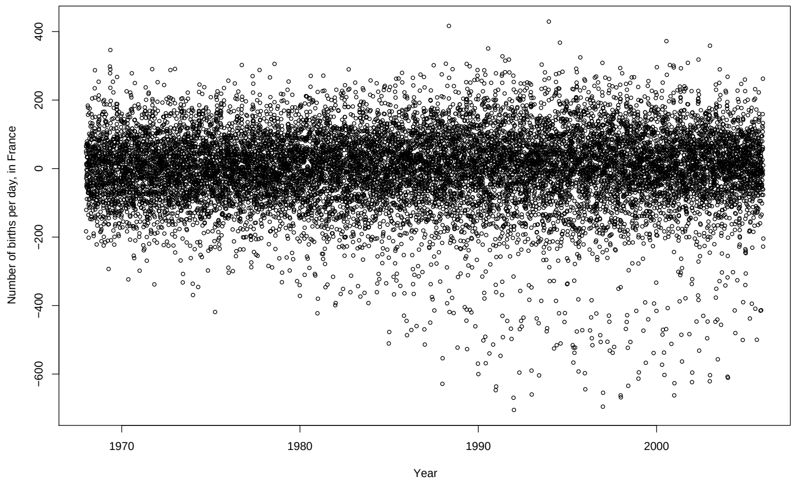

Figure 4: Residual noise of the series, corrected from the weekly effet, and when the general pattern is removed.

outliers are pointed out (1.96 %) and replaced by the mean of the four nearest data points (see Figure 5).

3 Full-moon effect

The variation of the average number of births along the distance to the full Moon (in particular 0 and 1 being the full Moon, 1/2 the new Moon, etc.) is now studied on the basis of the detrended series. Figure 6 shows the average on the overall data set (almost 40 years), and a comparison with the period 1968-1975 (dotted line) and the period 1999-2005 (plain line).

Let $\mathcal{S}_d(\lambda)$ denote the λ-days period beginning on a day at distance d of the next full Moon day. Analogously, $\bar{\mathcal{S}}_d(\lambda)$ stands for the $30-\lambda$ remaining days in the cycle (or $29-\lambda$ depending on the next full Moon day). For example, $\mathcal{S}_3(5)$ is a 5-days period which 3rd day is the full Moon day and $\mathcal{S}_{15}(6)$ is a 6-days period ending 10 days before the full Moon day.

We propose a multiple hypothesis testing methodology for the comparison between the mean number of daily births within $\mathcal{S}_d(\lambda)$ and $\bar{\mathcal{S}}_d(\lambda)$. Let $Z_{ij}^{(k)}$ denote the number of births at the jth day ($j = 1, \ldots, n_i$) of the ith period ($i = 1$ for days in $\mathcal{S}_d(\lambda)$ and 2 in $\bar{\mathcal{S}}_d(\lambda)$) in the kth lunar cycle. Hereafter, we focus on the simultaneous tests of $\mathrm{H}_0^{(k)}$: $\mathbb{E}(Z_{1j}^{(k)}) = \mu_{1k} = \mathbb{E}(Z_{2j}^{(k)}) = \mu_{2k}$ against $\mathrm{H}_1^{(k)}$: $\mu_{1k} > \mu_{2k}$, for $k = 1, \ldots, m$, where $m = 469$ is the number of lunar cycles in the dataset. In the following, the numbers of cycles with significant period effect are compared for all the possible distance d to the full Moon day and a fixed length of 6 days ($n_1 = 6$). The question of an optimal λ which highlights the full Moon effect is also

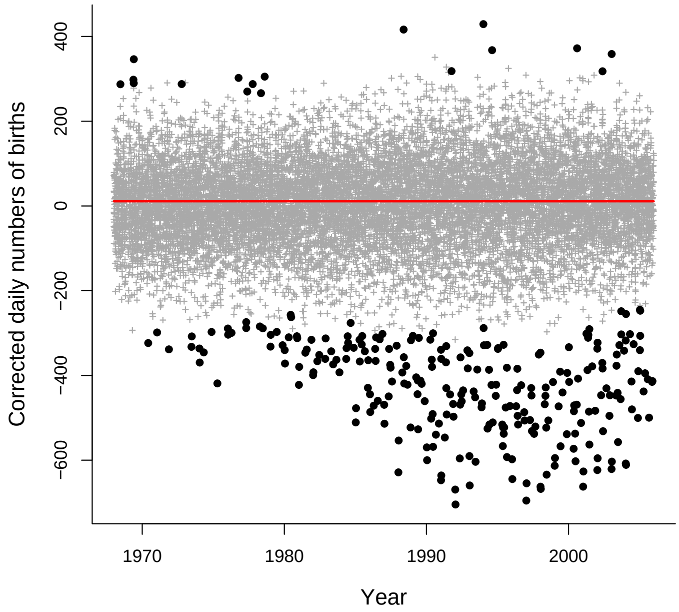

Figure 5: Outliers detection. The red curve represents the fit by a robust local regression method and the outliers (1.96 % of the data points) are marked by plain circles.

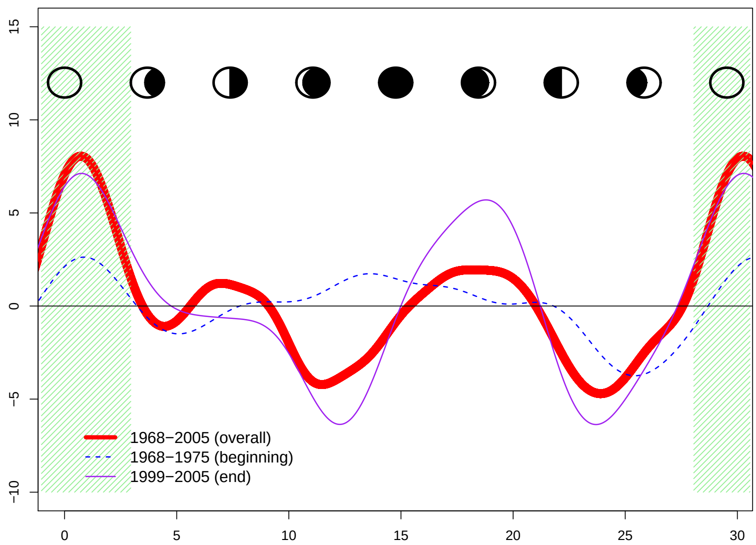

Figure 6: Average value for the mean detrended daily number of births along the moon cycle (in days).

addressed hereafter.

Multiple testing procedure

Among the most recent ideas to control the negative impact of dependence in multiple testing, [13] and [7] have both suggested to account for dependence between the test statistics by means of a factor modelling of the intra-group variance. Such a model assumes the existence of latent factors $F = (F^{(1)}, \ldots, F^{(q)})$, supposed to concentrate in a small dimension space the common information contained in the m responses: for $k = 1, \ldots, m$,

$$Z_{ij}^{(k)} \quad = \quad \mu + \alpha_i + b_k' F + \varepsilon_{ij}^{(k)}, \tag{1}$$

where b_k is the kth vector of loadings and the error terms $\varepsilon_{ij}^{(k)}$ are mutually independent, also independent of the factors, normally distributed with mean 0 and variance ψ_k^2, known as the kth specific variance. Moreover, it will be assumed that F is normally distributed with mean 0 and variance I_q, which is sometimes referred to as the exploratory factor analysis model. [7] show that, under assumption (1), asymptotically optimal test statistics are given by:

$$T^{(k)}(Z) \quad = \quad \frac{\bar{Z}_1^{(k)} - \bar{Z}_2^{(k)} - \hat{b}_k'(\bar{\hat{F}}_1 - \bar{\hat{F}}_2)}{\hat{\psi}_k \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{2}$$

where $\hat{b}_k$ and $\hat{\psi}_k$ are consistent estimators of the factor model parameters, $\bar{Z}_i^{(k)}$ and $\bar{\hat{F}}_i$ are respectively the means of $Z^{(k)}$ and of the $q-$vector of estimated scores $\hat{F}$ in the ith group. Under the null hypothesis, the above test statistics are independent and approximately distributed, in small-sample conditions, by a Student distribution which degrees of freedom, given in [7], are adjusted for the complexity of the factor model. In the sequel, an estimation procedure for the parameters of the factor analysis model is implemented, inspired by the EM approach discussed in [25].

Extraction of factors

Many estimation methods can be used in the factor analysis model, among which Principal Factoring is probably the most famous (see [14]). However, in high-dimensional situations, Principal Factoring can be computationally cumbersome because each step of the iterative algorithm consists in a singular value decomposition (SVD) of a large correlation matrix. Since factor analysis is a particular latent variable model, an EM algorithm (see [25]) can be implemented to achieve the maximum likelihood solution and avoid SVD of large matrices. The algorithm proposed by [7] and implemented hereafter transposes the initial EM algorithm to the multiple testing situation.

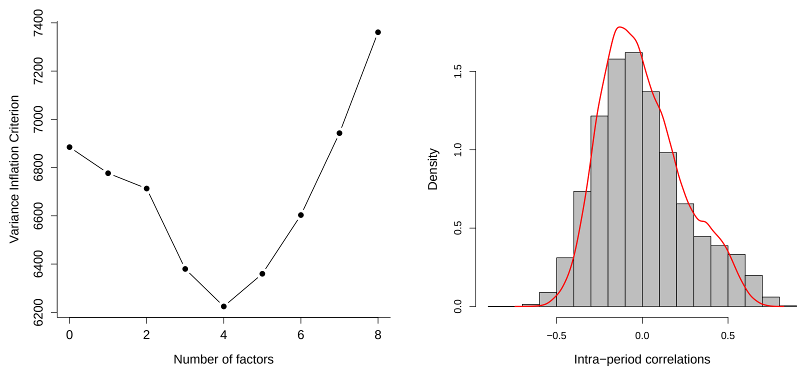

Figure 7: Left panel plot: variance inflation criterion along with the number of factors. Right panel plot: histogram of the intra-period correlation with the density curve of the fitted correlation model (4-factors model).

The first step is to properly estimate the number q of factors. In their study of the impact of dependence on the variance of the number V_τ of false rejections, [7] propose to estimate q by minimization of an ad-hoc criterion, which can be viewed as the amount of variance inflation due to the correlation between the test statistics. If, for a given number p of factors, R_p stands for the correlation matrix between the factor-adjusted test statistics (2), [7] show that this variance can be expressed as $\left[m_0 + M_\tau(R_p)\right]\tau(1-\tau)$, where $M_\tau(R_p)$ is the sum of U-shaped functions of all pairwise correlations in R_p and m_0 is the number of truly null hypotheses. As an illustration, if we consider the 6-days period beginning 3 days before the full Moon day, the right-panel plot of Figure 3, which displays the values of $\hat{M}_\tau$, for $\tau = 0.05$, along with the number of factors, shows a minimal variance inflation with 4 factors. In the following, the factor-analytic procedure is therefore implemented with $q = 4$ factors.

Control of the False Discovery Rate

Basically, multiple testing procedures can be viewed as the sequence of a single-hypothesis testing method applied to each test and the choice of a threshold τ for the p-values, under which the null hypothesis is rejected. For each τ, let V_τ denote the number of erroneous rejections and R_τ the number of rejections. The thresholding procedure aims at controlling an overall type-I error rate at a given level α. For highly dimensional data, it is now quite commonly accepted that a reasonable choice of type-I error is the actual False Discovery Proportion $\text{FDP}_\tau = V_\tau/R_\tau$, namely the proportion of rejected hypotheses which are erroneously rejected. The expected FDP_τ, also called the False Discovery Rate and denoted FDR_τ, is defined by [3] as $\text{FDR}_\tau = \mathbb{E}(\text{FDP}_\tau \,|R_\tau > 0)$.

For a given type-I level α, the following method is also proposed by [3] to choose a threshold τ_α with $\mathrm{FDR}_{\tau_\alpha} \leq \alpha$: $\tau_\alpha = \max_\tau \Big\{ \tau \in [0,1],\ \widehat{\mathrm{FDR}}_\tau \leq \alpha \Big\}$, where $\widehat{\mathrm{FDR}}_\tau = m_0 \tau / R_\tau$ is an FDR estimate if m_0 is assumed to be known. Substituting m_0 by an accurate estimation results in a more precise control of the FDR (see for instance [12] for a review of estimation procedures). In the following, the BH procedure is applied on the factor-adjusted p-values to control the FDR at level $\alpha = 0.10$. As proposed by [29], an estimation of m_0 obtained by smoothing the distribution of the p-values is plugged in the BH procedure.

Full-Moon effect

The above multiple testing methodology is implemented for all the possible values of d. Figure 3 shows the numbers of cycles with significant period effect along with d. The horizontal axis has been centered on the full Moon day to highlight this period. This figure shows that the fraction of cycles with significant period effect is small whatever the first day d of the period, since it never exceeds 8 % of the 469 lunar cycles (about 1 per year on average). However, the largest fractions of such cycles are reached for periods around the full Moon day (4 times more than any other period), which strengthens the feeling of an abnormally large number of lunar cycles with a full Moon effect. Note also that the mean difference, among the significant cycles, between the daily numbers of births within the full Moon period and within the rest of the moon cycle is about 95, which can be seen as quite small regarding the mean number of births per day in France (about 2.000).

The same pattern as shown on figure 3 is also observed for the other values of λ between 3 and 10. However, for $\lambda = 6$ days, the contrasts between the fraction of cycles with significant full Moon effect regarding any other period is the most markedly different. Finally, Figure 3 reproduces the same plots as Figure 3 but restricted either to 1968-1977 or to 1996-2005. It confirms the general feeling, already mentioned as a comment of Figure 6 of a more obvious full Moon effect in the most recent years than in 1968-1977.

4 Discussion

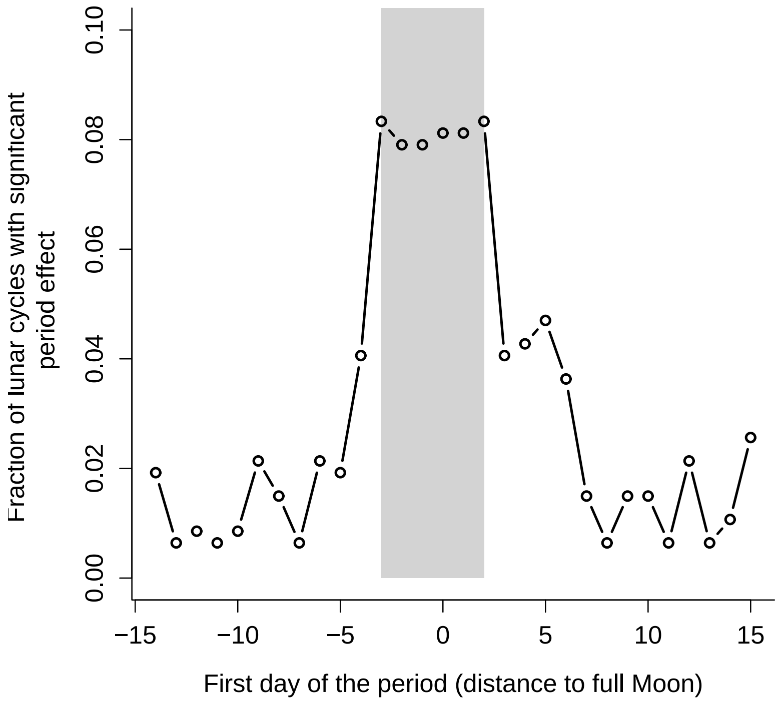

Figure 8: Fraction of cycles with significant period effect along the first day d of the period ($d = 0$ for the full Moon day). The 6-days period beginning 3 days before the full Moon is shaded to highlight the abnormally large fraction of cycles with full moon effect.

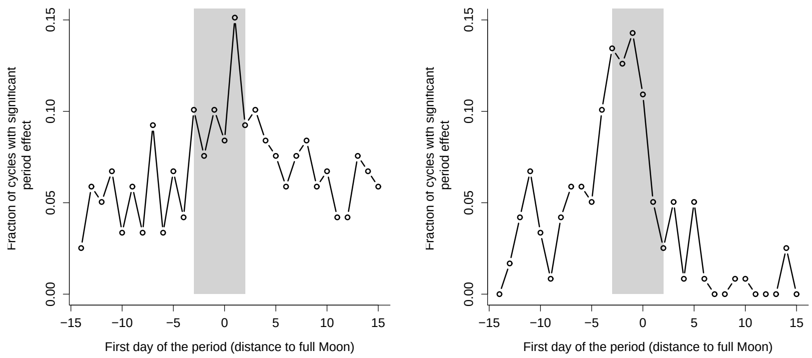

Figure 9: Fraction of cycles with significant period effect along the first day d of the period ($d = 0$ for the full Moon day). Left panel plot: 1968-1977. Right panel plot: 1996-2005

References

[1] Abell GO & Greenspan B. (1979). Human births and the phase of the Moon. *N Engl J Med.* 300, 96.

[2] Arliss J.M, Kaplan E.N. & Galvin S.L. The effect of the lunar cycle on frequency of births and birth complications. AM. J. Obstet. Gyncecol. 192, 1462-1464 (2005).

[3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society B*, 57, 289–300.

[4] Criss, T.B. & Marcum, J.P. A Lunar Effect on Fertility. Social Biology, 28, 75-80 (1981).

[5] Danzl D.F. (1987). Lunacy. *J. Emerg. Med.*, 5 91-95.

[6] Efron, B. (2007). Correlation and large-scale simultaneous significance testing, *J. Amer. Statist. Assoc.* 102, 93–103.

[7] Friguet, C., Kloareg, M. and Causeur, D. (2008). A factor model approach to multiple testing under dependence, *Submitted*.

[8] Ghiandoni G, Seclì R, Rocchi MB & Ugolini G. (1998). Does lunar position influence the time of delivery? A statistical analysis Eur J Obstet Gynecol Reprod Biol. 77, 47-50.

[9] Ghiandoni G, Seclì R, Rocchi MB, Ugolini G, & Cancelli V. (1998). Some unexpected results in time distribution analysis of spontaneous deliveries. *Gynecol Obstet Invest.* 46, 88-90.

[10] Guillon P, Guillon D, Lansac J, Soutoul JH, Bertrand P & Hornecker JP. Births, fertility, rhythms and lunar cycle. A statistical study of 5,927,978 births. J Gynecol Obstet Biol Reprod (Paris). 15, 265-271 (1986).

[11] Kelly IW & Martens R. Lunar phase and birthrate: an update. Psychol Rep. 75, 507-511 (1994)

[12] Langaas, M., Lindqvist, B.H. and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data, *Journal of the Royal Statistical Society. Series B*, 67, 216, 555–572.

[13] Leek, J.T. and Storey, J.D. (2008). A general framework for multiple testing dependence, *Proceedings of the National Academy of Sciences, USA*, 105, 18718–18723.

[14] Mardia, K. V., Kent, J. T., and Bibby, J. M.. Multivariate Analysis. Probability and Mathematical Statistics. Academic Press, London. (179)

[15] Martens R., Kelly I.W. & Saklofske D.H. Lunar phase and birthrate : a 50-year critical review. Psychol. Rep., 63 923-924 (1988).

[16] McDonald, RL. (1966) Lunar and seasonal variations in obstetric factors. *J Genet Psychol* 108 81-87.

[17] Meeusn J. Astronomical Algorithms. Willmann-Bell Publishers. (1991)

[18] Menaker W. (1967). Lunar periodicity with reference to live births. *Am J Obstet Gynecol.* 98, 1002-1004.

[19] Menaker, W. & Menaker, A. (1959). Lunar Periodicity in Human Reproduction: A Likely Unit of Biological Time. *Am J Obstet Gynecol*, 77, 905-914.

[20] Nalepka, CD., Jones, SL & Jones PK. Time variations, births, and lunar association, Issues Compr Pediatr Nurs 6, 81-89. (1983)

[21] Osley, M., Summerville, D., & Borst, L.B. (1973). Natality and the Moon. *Am J Obstet Gynecol* 117, 413-415.

[22] Periti E, Biagiotti R. Lunar phases and incidence of spontaneous deliveries. Our experience. Minerva Ginecol. 46, 429-433 (1994).

[23] Romero Martnez J, Guerrero Guijo I & Artura Serrano A. (2004). The Moon and delivery. *Rev Enferm.* 27, 11-12.

[24] Rousseeuw P.J. and Leroy A.M. Robust Regression and Outlier Detection. Wiley. (1987)

[25] Rubin, D.B. and Thayer, D.T. (1982). EM algorithms for ML factor analysis, *Psychometrika*, 47, 1, 69–76.

[26] Schnurman, A.G (1949). The effect of the Moon on childbirth. *Va Med Mon* 76-78.

[27] Shulman, H.B. Investigating lunar cycles in monthly fertility rates. SRD Research Report CENSUS/SRD/RR-87/12 (1987)

[28] Simonoff JS. (1996) Smoothing Methods in Statistics. Springer Verlag.

[29] Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies, *Proc. Nat. Acad. Sci. USA*, 100, 9440–9445.

[30] Strolego F, Gigli C, Bugalho A. The influence of lunar phases on the frequency of deliveries Minerva Ginecol. 43, 359-63 (1991)

[31] Ytterstad E. & Brenn T. (1997). Daily number of births in Norway 1989-93. Variations across months, day of week, phase of Moon and changes in leave entitlements. *Tidsskr. Nor Laegeforen*, 117 1098-1101.